Climate
of the Past

Open Access

Discussions

Interactive comment on "On reconstruction of time series in climatology" *by* V. Privalsky and A. Gluhovsky

Anonymous Referee #1

Received and published: 5 November 2015

When stripped down to its core elements, the current study presents a vector autoregressive model of order 3 to simulate the bivariate series of sunspot numbers (SSN) and total solar irradiance (TSI). The model is estimated from overlapping observations of both, and is meant to be applied to past SSN observations for obtaining past values of TSI.

Unfortunately, the paper is overloaded with a very large amount of jargon, almost assuming a bird's-eye view on climate research as a whole. For example, the abstract starts as follows:

"The approach to time series reconstruction in climatology based upon cross-correlation coefficients and regression equations is mathematically incorrect because it ignores the dependence of time series upon their past."

This view leads, consequently, to numerous statements about statistical methodology that are either simply wrong or completely trivial (see details below). Besides that, crucial elements are missing, such as a discussion of relevant literature (on SSN and TSI modeling in particular, not on time series modeling in general), a concise mathematical presentation of the core model, and its thorough verification on independent data. Other parts are unnecessary, such as repeating time domain results for the frequency domain. I therefore reject the paper for publication in CP.

The suggested model itself, if I understand correctly and adjust correspondingly, comes down to the equation

TSI(t) = a*SSN(t) + b*SSN(t-1) + c*SSN(t-2)

so that compared to "normal" regression on SSN, SSN autocorrelation is factored in. If the manuscript is stripped of all unnecessary jargon and mathematical formalism, so that this core is revealed, the approach may be salvaged and eventually become publishable. In passing, I note that another recent submission for ESD by the first author contains very similar material. A possible re-submission, therefore, would not only require the requested modificatinons (remove jargon and provide thorough literature review and proper verification) but would also have to be sufficiently distinguished from that study.

Some general remarks:

Most parts of section 3 are expendable. They contain nothing but standard material of timeseries analysis, and they are repetitive in displaying the 2-dim case and the n-dim case separately.

The validation R^2 scores are inadequate. The model is based on a 2-dim. vector autoregressive process for (SSN, TSI) whose predictions involve TSI memory. But that memory is not available when applied to reconstructions of TSI.

One may at least discuss, perhaps also consider, simulations based on SSN and TSI

anomalies from a harmonic base model.

Specific comments:

4702, 3: What means "mathematically" in this context?

4703, 12: statements like this one, "... regression approach is generally not correct." should be avoided, since it is unclear what exactly they refer to, and they do not serve any purpose other than distinguishing a study's work from others.

4705, 21: I cannot see any Bayesian approach in von Storch et al. (2004).

4706: I don't think the fact that correlation can be seen as coherence averaged over frequency deserves an entire page. Moreover, I don't think the equivalent formulation of correlation or coherence in terms of information content provides any additional insight.

4707: Please do not reinvent the wheel. Just mention that you employ a 2-dimensional AR(p) process.

4710, 15: Demanding correlations to be perfect means to give up on *any* attempt for building empirical statistical models. With the same argument, models based on coherence values below 1 could never be justified.

4711, 14: Optimal AR orders are based on which criterion?

, 23: In regression, explained variance and (squared) correlation are equivalent measures, so only one should be used.

4712, 1: The cross-correlation function does not look healthy. What causes the jump at lag zero?

, 8: time series can just as well be random variables, so that distinction is mute (see also similar phrases earlier in the manuscript).

, 13ff: should be removed as being far too general and unrelated to the subject

4713, 5: How does this relate to the orders of p=32/33 found earlier?

, 24ff: Because Eq. (11) is essentially contained in Eq. (12) (except for one time lag) this is trivial. This reinforces the need for validation against independent data.

4714, 7: The provided R^2 values are misleading when used to validate the predictive power of the two models. They contain for the AR(3) model the predictive power that comes from the lagged information of TSI itself, but that is to be predicted, so it must not be used. A real validation needs independent verification data.

4715, 11: How do you define the first value of TSI for 1749? Moreover, the lag-1 x_1 predictor has no extra predictive power since it is also being predicted. It can and must be dropped.

4716, 6ff: This needs to be estimated from out-of-sample data, because otherwise it is trivial (see above) as more parameters are available for the fit.

4717, 1ff: This has nothing to do with the subject of the study.

―――――――――――――――――――

Interactive comment on Clim. Past Discuss., 11, 4701, 2015.